What does Web 2.0 have to do with databases?

Sihem Amer-Yahia
Yahoo! Research
New York, NY
http://research.yahoo.com/sihem

Alon Halevy
Google
Mountain View, CA
http://alonhalevy.googlepages.com/

1. IMPORTANCE/RELEVANCE OF PANEL

Web 2.0 is a buzzword we have been hearing for over 2 years. According to Wikipedia,[1] it hints at an improved form of the World Wide Web where technologies such as weblogs, social bookmarking, RSS feeds, photo and video sharing, based on *an architecture of participation and democracy* that encourages users to add value to the application as they use it. Web 2.0 enables social networking on the Web by allowing users to contribute content[2], share it,[3] rate it,[4] create a network of friends,[5] and decide what they like to see and how they want it to look like.[6]

In this panel, we propose to explore what Web 2.0 has to do with databases. Among the questions we will ask are: *Are we going to be building new kinds of databases with Web 2.0? What does it mean when 1 million people are building a database as a collaborative effort? What is a new notion of schemas and queries that we need to develop for such scenarios?* The main point of controversy is that while some people think that Web 2.0 requires a new data management infrastructure, others view it as a great opportunity to integrate many "small" distributed databases which store users and content. We finish the panel by asking the question: *Are there databases we can build now that could not be built earlier because we have mass collaboration?*

1.1 Questions/Points of Controversy

Users: User has multiple identities. Is "user" the right granularity or should a user profile, which may represent a set of users, be? Databases do not commonly capture the notion of user. Can databases be extended to represent users? Do we need multiple databases to capture multiple profiles? Is this information stored or inferred?

Content: User-contributed content may be raw text, in the case of reviews, points, in the case where users rate other content, structured text, in the case of emails, photos, videos. Clearly databases as we know them cannot handle this wide-variety of data types. Is this a good reason not to consider using databases for Web 2.0 storage?

Search: The greatest thing about keyword search is its simplicity. How can search interfaces remain simple while leveraging *shared behavior* between users. Some think that users may be allowed to express search preferences. [7] Others think that they should contribute search algorithms.[8] But what are users searching for in the first place? the metadata (tags, reviews, ratings, comments), the social network (other people in the network), the content, all of the above? Gven all that, can we just do keyword search over databases? In fact, do we need to use databases at all?

Information rendering: What do users see when they come to a Webpage or as a result of a search? Hotlists are an example of information displayed to the user based on an "implicit" search. Beside the need for different layouts enabled by mashups (e.g., MAFIA), the result of a search may contained additional information such as the overlay of how many of the user's contacts viewed or saved some URL in MyWeb).

2. PANEL ORGANIZERS

- **Sihem Amer-Yahia** is a Senior Research Scientist at Yahoo! Research. She is interested in the interplay between structured data and search in online communities. Previously, she worked on XML search. In particular, she represented AT&T Labs on the Full-Text Task Force within the W3C. Sihem co-chaired WebDB 2004 and XSym 2006. Recently, she gave a keynote at WebDB 2007 on Web 2.0 search challenges.
- **Alon Halevy** is research scientist at Google, and previously a professor at the University of Washington. Alon's current research interests are focused on the concept of dataspaces, which attempt to bring data management techniques to wider audiences. Alon served as SIGMOD PC Chair in 2003 and has given several keynotes in recent years, including VLDB 2004 and PODS 2006.

3. PANELISTS

The list of confirmed panelists is: **AnHai Doan**, University of Wisconsin; **Anant Jhingran**, IBM; **Donald Kossmann**, ETH Zurich; **Gerhard Weikum**, Max-Planck Institute.

[1] *www.wikipedia.com*
[2] *www.wikipedia.com*
[3] *www.flickr.com*
[4] *www.youtube.com*
[5] *www.del.icio.us*
[6] *www.myspace.com*
[7] *www.myweb.com*
[8] *www.wikia.com*

VLDB '07, September 23-28, 2007, Vienna, Austria.